Conceptual inadequacy of the Shannon information in quantum measurements

Časlav Brukner and Anton Zeilinger

Institute for Experimentalphysics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria

In a classical measurement the Shannon information is a natural measure of our ignorance about properties of a system. There, observation removes that ignorance in revealing properties of the system which can be considered to preexist prior to and independent of observation. Because of the completely different root of a quantum measurement as compared to a classical measurement, conceptual difficulties arise when we try to define the information gain in a quantum measurement using the notion of Shannon information. The reason is that, in contrast to classical measurements, quantum measurements, with very few exceptions, cannot be claimed to reveal a property of the individual quantum system existing before the measurement is performed.

I. INTRODUCTION

In classical physics information is represented as a binary sequence, i.e., a sequence of bit values, each of which can be either 1 or 0. When we read out information that is carried by a classical system we reveal a certain bit value that exists even before the reading of information is performed. For example, when we read out a bit value encoded as a pit on a compact disk, we reveal a property of the disk existing before the reading process.

This means that in a classical measurement the particular sequence of bit values obtained can be considered to be physically defined by the properties of the classical system measured.[1] The information read is then measured by the Shannon measure of information [1] which can operationally be defined as the number of binary questions (questions with ''yes'' or ''no'' answers only) needed to determine the actual sequence of 0's and 1's.

In quantum physics information is represented by a sequence of qubits, each of which is defined in a two-dimensional Hilbert space. If we read out the information carried by the qubit, we have to project the state of the qubit onto the measurement basis $\{|0\rangle, |1\rangle\}$ which will give us a bit value of either 0 or 1. Only in the exceptional case of the qubit in an eigenstate of the measurement apparatus does the bit value observed reveal a property already carried by the qubit. Yet in general the value obtained by the measurement has an element of irreducible randomness and therefore cannot be assumed to reveal the bit value or even a hidden property of the system existing before the measurement is performed.

This implies that in a sequence of measurements on qubits in a superposition state $a|0\rangle + b|1\rangle$ ($|a|, |b| \neq \{0,1\}$) the particular sequence of bit values 0 and 1 obtained cannot, not

even in principle[2] be considered in any way to be defined before the measurements are performed. The nonexistence of well-defined bit values prior to and independent of observation suggests that the Shannon measure, as defined by the number of binary questions needed to determine the particular *observed* sequence 0's and 1's, becomes problematic and even untenable in defining our uncertainty as given *before* the measurements are performed.

Here we will critically analyze the applicability of the axiomatic derivation of the Shannon measure for the case of quantum measurement. We will also show that Shannon information is not useful in defining the information content in a quantum system. In fact we will see that when we try to apply Shannon's postulate in quantum measurements or when we try to define the information content by the Shannon information a certain element emerges that escapes complete and full description in quantum mechanics. This element is always associated with the objective randomness of individual quantum events and with quantum complementarily. In the end we will briefly discuss a novel and more suitable measure of information [4]. Yet at first we will re-

---

[1] Even in these cases where classical physics instead of definite measurement results predicts these results with certain probabilities, it is still possible, at least in principle, to consider an ensemble of statistically distributed measurement results as revealing corresponding statistically distributed properties of the ensemble of classical systems.

---

[2] As theorems like those of Kochen and Specker [2] show, it is fundamentally not possible to assign to a quantum system (noncontextual) properties corresponding to all possible measurements. The theorems assert that for a quantum system described in a Hilbert space of dimension equal to or larger than 3, it is possible to find a set of n projection operators which represent the yes-no questions about an individual system, such that none of the 2^n possible sets of answers is compatible with the sum rule of quantum mechanics for orthogonal decomposition of identity [3] (i.e., if the sum of a subset of mutually commuting projection operators is the identity one and only one of the corresponding answers ought to be ''yes''). This means that it is not possible to assign a definite unique answer to every single yes-no question represented by a projection operator independent of which subset of mutually commuting projection operators one might consider it with together. If there are no definite (context-independent) answers to all possible yes-no questions that can be asked about the system, then the operational concept of the Shannon measure of information itself, defined as the number of yes-no questions needed to determine the particular answers the system gives, becomes highly problematic.

---

turn to a discussion in more detail of the operational definition of Shannon information to quantum measurements.

II. DISCUSSION OF THE OPERATIONAL DEFINITION FOR A SEQUENCE OF MEASUREMENTS

For classical observations Shannon's measure of information can conceptually be motivated through an operational approach to the question. Consider an urn filled with N colored balls. There are $n_1, n_2, \ldots, n_m$ balls with various different colors: black, white, . . . , red. Now the urn is shaken, and we draw one after the other all balls from the urn. To what extent can we predict the particular color sequence drawn?

Certainly, if all the balls in the urn are of the same color, we can completely predict the color sequence. On the other hand, if the various colors are present in equal proportions and if we have no knowledge about the arrangement of the balls after shaking the urn, we are maximally uncertain about the color sequence drawn. One can think of these situations as extreme cases on a varying scale of predictability. For example, for $N=4$, there is only one color sequence $\bigcirc\bigcirc\bigcirc\bigcirc$ if all balls are white and four possible color sequences $\bullet\bigcirc\bigcirc\bigcirc$, $\bigcirc\bullet\bigcirc\bigcirc$, $\bigcirc\bigcirc\bullet\bigcirc$, $\bigcirc\bigcirc\bigcirc\bullet$, if there are three black and one white ball in the urn, yet six possible color sequences $\bullet\bullet\bigcirc\bigcirc$, $\bullet\bigcirc\bullet\bigcirc$, $\bullet\bigcirc\bigcirc\bullet$, $\bigcirc\bullet\bullet\bigcirc$, $\bigcirc\bullet\bigcirc\bullet$, $\bigcirc\bigcirc\bullet\bullet$ if there are two black and two white balls in the urn. This suggests that the uncertainty we have before drawing about the particular color sequence that will be drawn is defined by the total number of different possible color sequences that are in accordance with the given number of balls with their respective colors in the urn.

The total number of different color sequences can be obtained as the number of distinguishable permutations of N balls made up of m groups of black, white, . . . red balls indistinguishable within each group:

$$W = \frac{N!}{n_1! n_2! \cdots n_m!}.$$

For N sufficiently large,

$$W \simeq 2^{NH}, \tag{1}$$

where

$$H = -\sum_{i=1}^{n} p_i \log p_i \tag{2}$$

is the Shannon information expressed in bits when the logarithm is taken to base 2 and $p_1 = n_1/N$, $p_2 = n_2/N$, . . . ,$p_m = n_m/N$ are the proportions of balls with different colors in the urn.

Suppose now that one wishes to identify a specific color sequence of the drawn balls from the complete set of possible color sequences by asking questions to which only "yes" or "no" can be given as an answer. Of course, the number of questions needed will depend on the questioning strategy adopted. In order to make this strategy the most optimal, that is, in order that we can expect to gain maximal
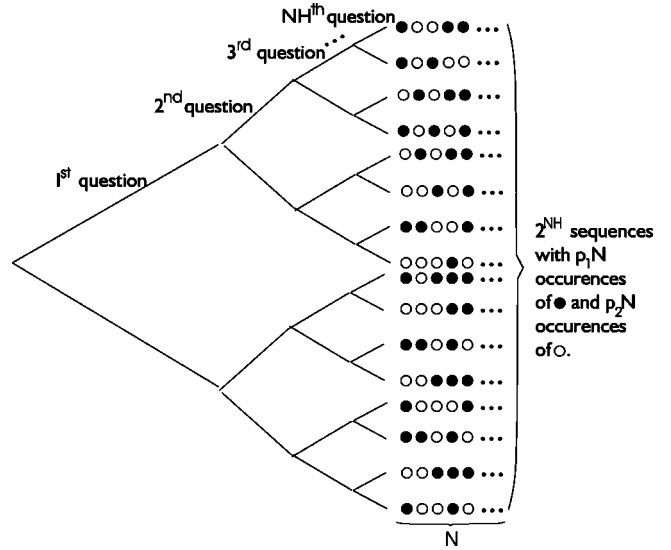

FIG. 1. Binary question tree to determine the specific sequence of outcomes (color of the drawn balls) in a sufficiently large number N of experimental trials (number of drawings). An urn is filled with black and white balls in proportions p_1 and p_2, respectively. The expected number of questions needed to determine the actual sequence of outcomes is NH, where $H = -p_1 \log p_1 - p_2 \log p_2$.

information from each yes-or-no question, we evidently have to ask questions whose answers will strike out always half of the possibilities.

Since there are $W = 2^{NH}$ possible different color sequences (all of them have equal probability to be drawn), the minimal number of yes-no questions needed is just NH. Or, equivalently, the Shannon information expressed in bits is the minimal number of yes-no questions necessary to determine which particular sequence of outcomes occurs, divided by N [5–7]. A particular color sequence is specified by writing down, in order, the yes's and no's encountered in traveling from the root to the specific leaf of the tree as schematically depicted in Fig. 1 for an explicit example with an urn containing black and white balls only.

If instead of balls with preassigned colors we consider quantum systems whose individual properties are not defined before the measurements are performed, does the Shannon measure of information still define the information gain in the measurements appropriately? More precisely, we ask here the question whether the total number $W = 2^{NH}$ of different possible sequences of outcomes is suitable as a measure of our uncertainty before the sequence of quantum measurements is performed.

In classical physics the behavior of the whole ensemble follows from the behavior of its intrinsically different individual constituents which can be thought to be defined to any precision. This is not the case in quantum mechanics. The indefiniteness in principle in the sense of the fundamental nonexistence of a detailed description of and prediction for the individual quantum event resulting in a specific measurement result, implies that the particular sequence of outcomes specified by writing down, in order, the yes's and no's encountered in a row of yes-no questions asked is not defined before the measurements are performed. No definite out-
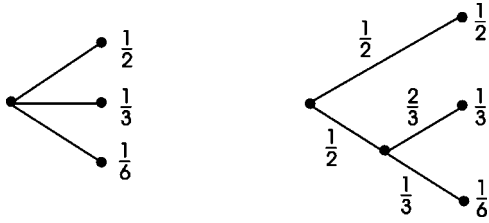
FIG. 2. Decomposition of a choice from three possibilities. Figure taken from [1].

comes exists before measurements are performed and therefore the number of different possible sequence of outcomes does not characterize our uncertainty about the individual system before measurements are performed.

However, once the sequence of quantum measurements is performed and the measurement results are obtained, the measure of information needed to specify the particular sequence of outcomes realized is defined appropriately by the Shannon measure. In the sense that an individual quantum event manifests itself only in the measurement process and is not precisely defined before the measurement is performed, we may speak of ''generation'' of that specific information in the measurement.

III. INAPPLICABILITY OF SHANNON'S POSTULATES IN QUANTUM MEASUREMENTS

An important reason for preferring the Shannon measure of information lies in the fact that it is uniquely characterized by Shannon's intuitively reasonable postulates. This has been expressed strongly by Jaynes [8]: ''One . . . important reason for preferring the Shannon measure is that it is the only one that satisfies . . . [Shannon's postulates]. Therefore one expects that any deduction made from other information measures, if carried far enough, will eventually lead to contradiction.'' A good way to continue our discussion is by reviewing how Shannon, using his postulates, arrived at his famous expression. He writes the following [1]

''Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much 'choice' is involved in the selection of the event or how uncertain we are of the outcome?

If there is such a measure, say, $H(p_1, p_2, \ldots, p_n)$, it is reasonable to require of it the following properties:

1. H should be continuous in the p_i.

2. If all the p_i are equal, $p_i = 1/n$, then H should be a monotonically increasing function of n. With equally likely events there is more choice, or uncertainty, when there are more possible events.

3. If a choice is broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated in Fig. 2. At the left we have three possibilities $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, and $p_3 = \frac{1}{6}$. On the right we first choose between two possibilities each with probability $\frac{1}{2}$ and, if the second occurs, make another choice with probabilities $\frac{2}{3}, \frac{1}{3}$. The final results have

the same probabilities as before. We require, in this special case, that

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}H\left(\frac{2}{3}, \frac{1}{3}\right).$$

The coefficient $\frac{1}{2}$ is the weighing factor introduced because this second choice occurs half the time.''

Shannon then shows that only the function (2) satisfies all three postulates. It is the third postulate which determines the logarithm form of the function and, as we will argue, it is this postulate which leads to problems when quantum measurements are involved.

We now turn to the discussion of Shannon's postulates. While the first two postulates are natural for every meaningful measure of information, the last postulate might deserve more justification. The third Shannon postulate originally formulated as an example was reformulated as an exact rule by Faddeev [9]: For every $n \geq 2$,

$$H(p_1, \ldots, p_{n-1}, q_1, q_2) = H(p_1, \ldots, p_{n-1}, p_n)$$

$$+ p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right), \quad (3)$$

where $p_n = q_1 + q_2$.

Without physical interpretation the recursion postulate (3) is merely a mathematical expression which is certainly necessary for the uniqueness of the function (2) but has no immediate physical significance. We adopt the following well-known interpretation [6,10]. Assume the possible outcomes of the experiment to be $a_1, \ldots, a_n$ and $H(p_1, \ldots, p_n)$ to represent the amount of information that is gained by the performance of the experiment. Now, decompose event a_n into two distinct events $a_n \wedge b_1$ and $a_n \wedge b_2$ (''$\wedge$'' denotes ''and''; thus $a \wedge b$ denotes a joint event). Denote the probabilities of outcomes $a_n \wedge b_1$ and $a_n \wedge b_2$ by q_1 and q_2, respectively. Then the left-hand side $H(p_1, \ldots, p_{n-1}, q_1, q_2)$ of Eq. (3) represents the amount of information that is gained by the performance of the experiment with outcomes $a_1, \ldots, a_{n-1}, a_n \wedge b_1, a_n \wedge b_2$.

When the outcome a_n occurs, the conditional probabilities for b_1 and b_2 are q_1/p_n and q_2/p_n, respectively, and the amount of information gained by the performance of the conditional experiment is $H(q_1/p_n, q_2/p_n)$. Hence the recursion requirement states that the information gained in the experiment with outcomes $a_1, \ldots, a_{n-1}, a_n \wedge b_1, a_n \wedge b_2$ equals the *sum* of the information gained in the experiment with outcomes $a_1, \ldots, a_n$ and the information gained in the conditional experiment with outcomes b_1 or b_2, given that the outcome a_n occurred with probability p_n.

This interpretation implies that the third postulate can be rewritten as

$$H(p(a_1), \ldots, p(a_{n-1}), p(a_n \wedge b_1), p(a_n \wedge b_2))$$

$$= H(p(a_1), \ldots, p(a_{n-1}), p(a_n))$$

$$+ p(a_n)H(p(b_1|a_n), p(b_2|a_n)), \quad (4)$$

where

$$p(a_n) = p(a_n \wedge b_1) + p(a_n \wedge b_2),$$

$$p(a_n \wedge b_1) = p(a_n)p(b_1|a_n), \qquad (5)$$

$$p(a_n \wedge b_2) = p(a_n)p(b_2|a_n).$$

Here $p(b_i|a_n)$ $i=1,2$, denotes the conditional probability for outcome a_n given the outcome b_i occurs and $p(a_n \wedge b_i)$ denotes the joint probability that the outcome $a_n \wedge b_i$ occurs.

If we analyze the generalized situation with n outcomes a_i of the first experiment A, m outcomes b_j of the conditional experiment B, and mn outcomes $a_i \wedge b_j$ of the joint experiment $A \wedge B$, we may then rewrite the recursion postulate in short form as

$$H(A \wedge B) = H(A) + H(B|A), \qquad (6)$$

where $H(B|A) = \Sigma_j^n p(a_j) H(b_1|a_j, \ldots, b_m|a_j)$ is the average information gained by observation B given that the conditional outcome a_j occurred weighted by probability $p(a_j)$ for a_j to occur.

It is essential to note that the recursion postulate is inevitably related to the manner in which we gain information in a classical measurement. In fact, in classical measurements it is always possible to assign to a system simultaneously attributes corresponding to all possible measurements, here a_i, b_j and $a_i \wedge b_j$. Also, the interaction between measuring apparatus and classical system can be thought to be made arbitrarily small so that the experimental determination of A has no influence on our possibility to predict the outcomes of the possible future experiment B. In conclusion, the information expected in a classical experiment from the joint experiment $A \wedge B$ is simply the sum of the information expected from the first experiment A and the conditional information of the second experiment B with respect to the first, as expressed in Eq. (6).

Therefore, only for the special case of commuting, i.e., simultaneously definite observables, is the axiomatic derivation of the Shannon measure of information applicable and the use of the Shannon information justified to define the uncertainty given before quantum measurements are performed. However, in general, if A and B are noncommuting observables, the joint probabilities on the left-hand side of Eq. (4) cannot in principle be assigned to a system simultaneously, and consequently Shannon's crucial third postulate which is necessary for the uniqueness of Shannon's measure of information ceases to be well defined.

Having seen that the third Shannon postulate in general is not applicable in quantum measurements we next introduce two requirements that are immediate consequences of Shannon's postulates and in which all the probabilities that appear are well defined in quantum mechanics. We will show that the two requirements are violated by the information gained in quantum measurements, implying that the Shannon measure loses its preferential status with respect to alternative expressions when applied to define information gain in quantum measurements.

(1) *Every new observation reduces our ignorance and increases our knowledge.* In his work Shannon [1] offers a list of properties to substantiate that H is a reasonable measure of information. He writes, ''It is easily shown that

$$H(A \wedge B) \leq H(A) + H(B)$$

with equality only if the events are independent (i.e., $p(a_i \wedge b_j) = p(a_i)p(b_j)$). The uncertainty of a joint event is less than or equal to the sum of the individual uncertainties.'' He continues further in the text: '' ... we have

$$H(A) + H(B) \geq H(A \wedge B) = H(A) + H(B|A).$$

Hence,

$$H(B) \geq H(B|A). \qquad (7)$$

The uncertainty of B is never increased by knowledge of A. It will be decreased unless A and B are independent events, in which case it is not changed'' (we have changed Shannon's notation to coincide with that of our work).

(2) *Information is irrespective of the order of acquisition.* The total amount of information gained in successive measurements is independent of the order in which it is acquired, so that the amount of information gained by the observation of A followed by the observation of B is equivalent to the amount of information gained from the observation of B followed by the observation of A:

$$H(A) + H(B|A) = H(B) + H(A|B). \qquad (8)$$

This is an immediate consequence of the recursion postulate which can be obtained when we write the recursion postulate in two different ways depending on whether the observation of A is followed by the observation of B or vice versa. An explicit example for a sequence of classical measurements is given in Fig. 3.

Are these two requirements satisfied by the information gained in quantum measurements? Consider a beam of randomly polarized photons. Filters $F_\updownarrow$, $F_{45°}$, and $F_\leftrightarrow$ are oriented vertically, at $+45°$, and horizontally, respectively, and can be placed so as to intersect the beam of photons (Fig. 4). If we insert filter $F_\updownarrow$, the number of photons observed at the detection plate will be approximately half of the number in the incoming beam. The outgoing photons now all have vertical polarization. Notice that the function of filter $F_\updownarrow$ cannot be explained as a ''sieve'' that only lets those photons pass that are *already* of vertical polarization in the incoming beam. If that were the case, only a certain small number of the randomly polarized incoming photons would have vertical polarization, so we would expect a much larger attenuation of the beam of photons as they pass the filter.

Denote by A and B the properties of the photon to have polarization at $+45°$ and horizontal polarization, respectively. If $F_\leftrightarrow$ is inserted behind the filter $F_\updownarrow$, we are certain that none of the photons will pass through [Fig. 4(a)]. For a photon with vertical polarization we have complete knowledge of the property B, i.e., $H(B)=0$. Notice that a ''sieve'' model could explain this behavior. If we now insert $F_{45°}$ between $F_\updownarrow$ and $F_\leftrightarrow$, we observe an effect which cannot be explained by a sieve model where the filter does not change the object. In contrast, we now observe a certain number of
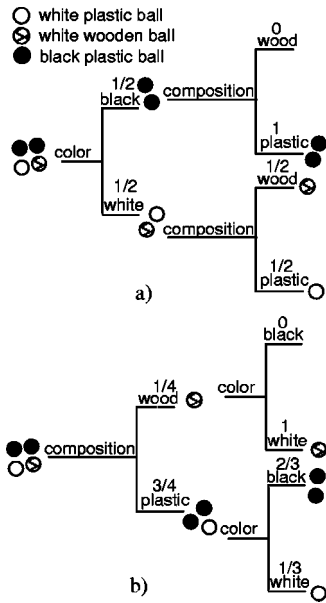
FIG. 3. Indifference of information to the order of its acquisition in classical measurements. A box is filled with balls of different compositions (plastic and wooden balls) and different colors (black and white balls). Then the box is shaken. In (a) we first draw a ball asking about the color of the drawn ball and gain $H(\text{color}) = 1$ bit of information. Subsequently, we put the black and white balls in separate boxes, draw a ball from each box separately, and ask about the composition of the drawn ball. We gain $H_{bl}(\text{comp}) = 0$ bits for the black balls and $H_{wh}(\text{comp}) = 1$ bit for the white balls. In (b) we pose the two questions in the opposite order. We first ask about the composition of the drawn ball and gain $H(\text{comp}) = 0.81$ bits. Then we ask about the color of the drawn ball and gain $H_{wo}(\text{color}) = 0$ bits for wooden balls and $H_{pl}(\text{color}) = 0.92$ bits for plastic balls. The total information gained is independent of the particular order the two questions are posed, i.e., $H(\text{color}) + 1/2H_{bl}(\text{comp}) + 1/2H_{wh}(\text{comp}) = H(\text{comp}) + 1/4H_{wo}(\text{color}) + 3/4H_{pl}(\text{color}) = 1.5$.

photons at the detection plate (about $\frac{1}{4}$ of the number of photons in the beam passed through $F_\uparrow$) as shown in Fig. 4(b). In this case our knowledge of the property B is not complete anymore.

The acquisition of information about property A therefore leads to a decrease of our knowledge about property B, i.e., $H(B|A) > 0$. Note that on the photons absorbed by the filter $F_{45°}$ we cannot measure property B subsequently. However, already for the subensemble of the photons passing through the filter our uncertainty about property B becomes larger than 0, implying $0 = H(B) < H(B|A)$, which clearly violates requirement (7). Another example of sequence of quantum measurements where requirement (8) is violated is shown in Fig. 5. Clearly, violation of the requirements (7) and (8) occurs when the corresponding operators A and B do not commute.

What is the origin of the violation of the requirements (7) and (8) in quantum measurements? In contrast to a classical measurement which just adds some new knowledge to our knowledge at hand from the previous measurements, in a quantum measurement the gain of the new knowledge is always at the expense of irrecoverable loss of complementary
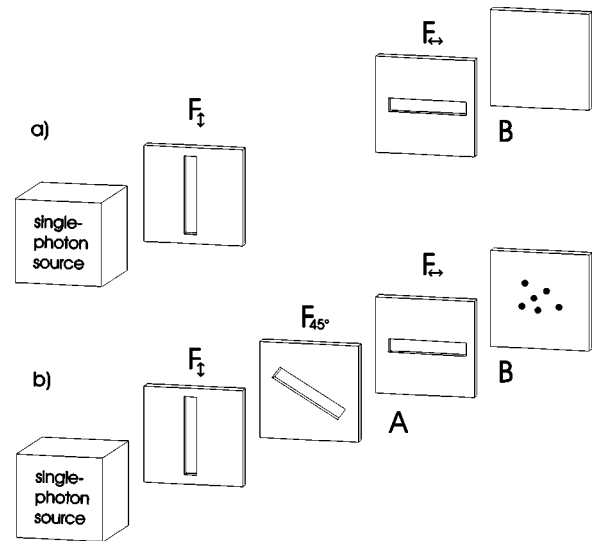

FIG. 4. The gain of knowledge in a new observation reduces our knowledge at hand from a previous observation. Filters $F_\uparrow$, $F_{45°}$, and $F_\leftrightarrow$ are oriented vertically, at $+45°$ and horizontally, respectively. If filter $F_\leftrightarrow$ is inserted behind the filter $F_\uparrow$, no photons are observed at the detector plate (a). In this case our knowledge about horizontal polarization (property B) of a photon passing through filter $F_\uparrow$ is complete. If filter $F_{45°}$ is inserted between $F_\uparrow$ and $F_\leftrightarrow$, a certain number of photons (1/4 of the number of photon passing through $F_\uparrow$) will be observed at the detection plate (b). Now acquisition of information about the polarization at $+45°$ (property A) leads to the decrease of our previous knowledge about horizontal polarization of the photon.

classes of knowledge. This originates from the distinction between ''total'' and ''complete'' information in quantum physics. In classical physics the total information about a system is complete. In quantum physics the total information of a system, represented by the state vector, is never complete in the sense that all possible future measurement results are precisely defined.[3] In fact, the total information of a quantum system suffices to specify the eigenstate of one nondegenerate (with one-dimensional eigenspaces only) observable only.

––––––––

[3] Yet we do not hesitate to emphasize that the information certainly is complete in the sense that it is not possible to have more information about a system than what can be specified in its quantum state. In fact, the state vector represents that part of our knowledge about the history of a system which is necessary to arrive at the maximum possible set of probabilistic predictions for all possible future observations of the system. For example, a set of complex amplitudes of a ψ-function is a specific representation of the catalog of our knowledge of the system. This view was assumed by Schrödinger [11] who wrote, ''Sie (die ψ-Funktion) ist jetzt das Instrument zur Voraussage der Wahrscheinlichkeit von Maßzahlen. In ihr ist die jeweils erreichte Summe theoretisch begründeter Zukunfterwartungen verkörpert, gleichsam wie in einem *Katalog* niedergelegt.'' Translated: ''It (the ψ function) is now the instrument for predicting the probability of measured numbers. In it is embodied the momentarily attained sum of theoretically based future expectations, somewhat as laid out in a catalog.''

For example, the state of a photon passing through filter $F_{\updownarrow}$ is fully specified by the complete knowledge about the property A of vertical polarization. If we let a photon in this state pass through filter $F_{45°}$ as given in Fig. 4(b), our knowledge of the photon changes, and therefore its representation, the quantum state, also changes. The total information of a photon in the new state is completely exhausted in specifying property B of polarization at 45° and no further information is left to also specify property A, thus implying unavoidable loss of the previous knowledge about this property. This further implies that the set of future probabilistic predictions specified by the new projected state is indifferent to the knowledge collected from the previous measurements in the whole history of the system. Such a view was assumed by Pauli [12] who writes,[4] ''Bei Unbestimmtheit einer Eigenschaft eines Systems bei einer bestimmten Anordnung (bei einem bestimmten Zustand des Systems) vernichtet jeder Versuch, die betreffende Eigenschaft zu messen, (mindestend teilweise) den Einfluß der früheren Kenntnisse vom System auf die (eventuell statistischen) Aussagen über spätere mögliche Messungsergebnisse.'' This clearly makes it possible to violate requirements (7) and (8) in quantum measurements.

Here a certain misconception might be put forward that arises from a certain practical point of view. According to that view, for example, complementarily between interference pattern and information about the path of the particle in the double-slit experiment is considered to arise from the fact that any attempt to observe the particle path would be associated with an uncontrollable disturbance of the particle. Such a disturbance in itself would then be the reason for the loss of the interference pattern. In such a view, it would be possible to define Shannon's information for all attributes of the system simultaneously, and the third Shannon postulate, as well as the requirements (7) and (8), would just be violated because of the unavoidable disturbance of the system occurring whenever the subsequently measured property B is incompatible with the previous one A. Yet this is a misconception for two reasons.

First, as theorems like those of Bell [13] or Greenberger, Horne, and Zeilinger [14] show, it is not possible, not even *in principle*, to assign to a quantum system simultaneously observation-independent properties which in order to be in agreement with special relativity have to be local. We therefore cannot speak of a ''disturbance'' in the measurement process if there are no objective properties to disturb.

Second, over the last few years experiments were considered and some already performed where the reason why no interference pattern arises is not due to any uncontrollable disturbance of the quantum system or the clumsiness of the apparatus. Rather the lack of interference is due to the fact that the quantum state is prepared in such a way so as to permit path *information* to be obtained, in principle, independent of whether the experimenter cares to read it out or not. One line of such research considers the use of micromasers in atomic beam experiments [15]; another one concerns experiments on correlated photon states emerging from nonlinear crystals through the process of parametric-down conversion [16].

The view that complementarity must be based on the much more fundamental property of mutual exclusiveness of different classes of information of a quantum system was emphasized by Pauli [12] in the analysis of the uncertainty relations:[5] '' . . . diese Relationen enthalten die Aussage, daß jede genaue Kenntnis des Teilchenortes zugleich eine prinzipielle Unbestimmtheit, nicht nur Unbekanntheit des Impulses zur Folge hat und umgekehrt. Die Unterscheidung zwischen (prinzipieller) *Unbestimmtheit* und *Unbekanntheit* und der Zusammenhang beider Begriffe sind für die ganze Quantentheorie entscheidend.''

IV. DIFFICULTIES IN DEFINING THE INFORMATION CONTENT OF A QUANTUM SYSTEM

To define the information content of a physical system one might consider different measures of information. However, only those measures of information have physical significance according to which the defined information content of the system possesses properties which naturally follow from the physical situation considered. These properties are, for example, invariance under changes of the modes of observation of the system and conservation in time if there is no information exchange with an environment. We show now that the information content of a quantum system, if it is assumed to be measured by the Shannon measure of information, cannot be defined in any way to have these properties.

The classical world appears to be composed of particles and fields, and the properties of each one of these constituents can be specified quite independently of the particular phenomenon discussed or of the experimental procedure a physicist chooses to determine these properties. In other words the properties of constituents of the classical world are noncontextual.

In particular, the total lack of information about a classical pointlike system (with no rotational and internal degrees of freedom) defined as Shannon's information associated with the probability distribution over the phase space is independent of the specific set of variables chosen to describe the system completely (such as position and momentum, or bijective functions of them) and conserved in time if there is no information exchange with an environment, i.e., if the system is dynamically independent of the environment and

---

[4]In translation: ''In the case of undefinedness of a property of a system for a certain arrangement (with a certain state of the system) any attempt to measure that specific property destroys (at least partially) the influence of earlier knowledge of the system on (possibly statistical) statements about later possible measurement results.''

---

[5]In translation:'' . . . these relations contain the statement that any precise knowledge of the position of a particle implies a fundamental undefinedness, not just an unknownness, of the momentum for a consequence and vice versa. The distinction between (fundamental) *indefiniteness* and *unknownness*, and the relation between these two notions is decisive for the whole quantum theory.''

not exposed to a measurement.[6] Operationally the total information content of a classical system can be obtained in the joint measurement of position and momentum or in successive measurements in which the observation of position is followed by the observation of momentum or vice versa.[7]

Contrary to the classical concepts most quantum-mechanical concepts are limited to the description of phenomena within some well-defined experimental context, that is, always restricted to a specific experimental procedure the physicist chooses. In particular the amount of information gained in an individual quantum measurement depends strongly on the specific experimental context. In the optimal experiment when the measurement basis $|i\rangle$ coincides with the eigenbasis of the density matrix $\hat{\rho}$ of the system: $\hat{\rho}|i\rangle = w_i|i\rangle$, the amount of information gained is minimal (see, for example, [3]). Since in the basis corresponding to the optimal experiment the density operator is represented by a diagonal matrix with elements w_i, the information gain defined by the Shannon measure equals the von Neumann entropy $s(\hat{\rho})$ as given by[8]

$$H = -\sum_i w_i \log w_i = -\mathrm{Tr}(\hat{\rho} \log \hat{\rho}) \equiv s(\hat{\rho}). \quad (11)$$

This has the important property of being invariant under unitary transformations $\hat{\rho} \rightarrow \hat{U}\hat{\rho}\hat{U}^+$. The invariance under unitary transformations implies invariance under the change of the representation (basis) of $\hat{\rho}$ and conservation in time if

---

[6]Given the probability distribution $\rho(\vec{r},\vec{p},t)$ over the phase space the total lack of information of a classical system is defined as [17]

$$H_{total}(t) = -\int d^3\vec{r}\,d^3\vec{p}\,\rho(\vec{r},\vec{p},t)\log\frac{\rho(\vec{r},\vec{p},t)}{\mu(\vec{r},\vec{p})}, \quad (9)$$

where a background measure $\mu(\vec{r},\vec{p})$ is an additional ingredient that has to be added to the formalism to ensure invariance under change of variables when we consider continuous probability distributions. The conservation of H_{total} in time for a system with no information exchange with an environment is implied by the Hamiltonian evolution of a point in phase space.

[7]In full analogy with Eq. (8) we may write $H_{total} = H(\vec{r}) + H(\vec{p}|\vec{r}) = H(\vec{p}) + H(\vec{r}|\vec{p})$.

[8]For a *given* density matrix $\hat{\rho}$ the von Neumann entropy

$$S(\hat{\rho}) = -\mathrm{Tr}(\hat{\rho} \log \hat{\rho}) \quad (10)$$

is widely accepted as a suitable definition for the information content of a quantum system. For a system described in N-dimensional Hilbert space this ranges from $\ln N$ for a completely mixed state to 0 for a pure state. The von Neumann entropy has the important property to be invariant under unitary transformations. However, we observe that any function of the form $\mathrm{Tr}[f(\hat{\rho})]$ [the operator $f(\hat{\rho})$ is identified by having the same eigenstates as $\hat{\rho}$ and the eigenvalues $f(w_j)$, equal to the function values taken at the eigenvalues w_j of $\hat{\rho}$] possesses this invariance property. We also observe that the von Neumann entropy is a property of the quantum state as a whole without explicit reference to information contained in individual measurements.
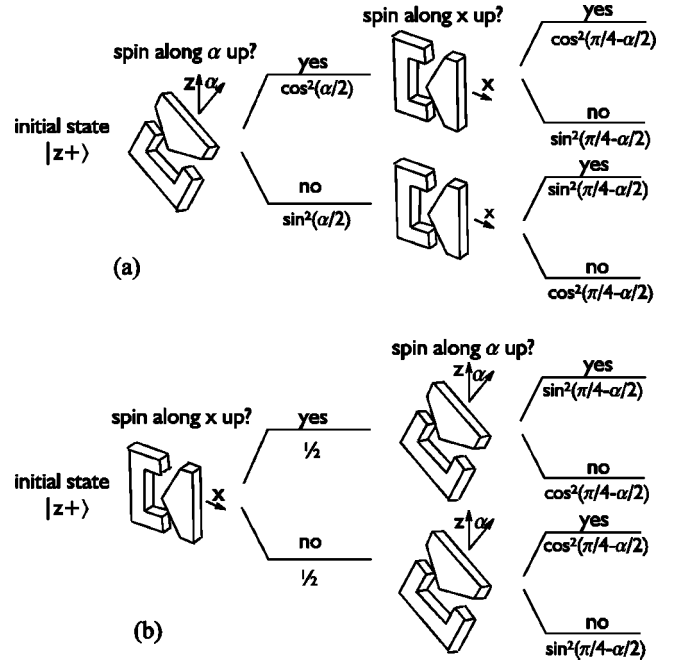

FIG. 5. Dependence of information on the order of its acquisition in successive quantum measurements. A spin-1/2 particle is in the state $|z+\rangle$ spin up along the z axis. Spin along the x axis and spin along the direction in the x-z plane tilted at an angle α from the z axes are successively measured, in one order in (a) and in the opposite order in (b). Whereas we obtain an equal portion $H(\cos^2(\pi/4 - \alpha/2), \sin^2(\pi/4 - \alpha/2))$ of information in the conditional (subsequent) measurement both in (a) and in (b), the amounts of information $H(\cos^2\alpha/2, \sin^2\alpha/2)$ and $H(\frac{1}{2}, \frac{1}{2}) = 1$ we gain in the first measurement in (a) and in the first measurement in (b), respectively, can be significantly different. Specifically for $\alpha \rightarrow 0$ we have complete knowledge about spin along the direction at the angle α in (a) but absolutely no knowledge about the spin along the x axis in (b). We emphasize that we do not assume any specific functional dependence for the measure of information H.

there is no information exchange with an environment. The latter precisely means that if we perform the optimal experiments both at time t_0 and at some future time t, Shannon's information measures associated with the optimal experiments at the two times will be the same, i.e.,

$$H(t) = -\sum_i w_i(t)\log w_i(t) = -\sum_i w_i \log w_i = H(t_0). \quad (12)$$

Here, the eigenvalues of the density matrix at time t are $w_i(t) = w_i$.

However, without the *additional* knowledge of the eigenbasis of the density matrix $\hat{\rho}$ we cannot find the optimal experiment and therefore we cannot obtain directly the Shannon information associated. Also, all the statistical predictions that can be made for the optimal measurement are the same as if we had an ordinary (classical) mixture, with fractions w_i of the systems each giving with certainty results that are associated with the eigenvectors $|i\rangle$. In this sense the optimal measurement is a classical type measurement and

therefore in this particular case, and only then, does Shannon's measure define the information gain in a measurement appropriately.[9] Considering also our previous discussion it is therefore not surprising that Shannon's measure is useful only when applied to measurements which can be understood as classical measurements.

Which set of individual measurements should we perform and how to combine individual measures of information obtained in the set in order to arrive at the information content of a quantum system if we do not know the eigenbasis of the density matrix? Quantum complementarity implies that the total information content of the system might be partially encoded in different mutually exclusive (complementary) observables. These have the property that complete knowledge of the eigenvalue of any one of the observables excludes *any* knowledge about the eigenvalues of all other observables. Such a set of observables for a spin-1/2 particle can for example be spin components along orthogonal directions.

We consider now a quantum system described in n-dimensional Hilbert space and we denote a complete set of m mutually complementary observables[10] by $\{\hat{A}, \hat{B}, \ldots\}$. The property of mutual expansiveness implies that if the system is in an eigenstate of one of the observables, for example, in the eigenstate $|a_j\rangle$ of the observable $\hat{A}$, and we measure any other observable from the set, say, $\hat{B}$, projecting the system onto states $\{|b_1\rangle, \ldots, |b_i\rangle, \ldots, |b_n\rangle\}$, the indi-

vidual outcome is completely random (all measurement results are equally probable):

$$|\langle a_j|b_i\rangle|^2 = \frac{1}{n} \quad \forall i,j. \quad (13)$$

It was shown in [19] that the density matrix of the system can fully be reconstructed if one performs a complete set of mutually complementary observations. This suggest that the total information content of a quantum system represented by a density matrix $\hat{\rho}$ is *all* obtainable from a complete set of mutually complementary measurements. To obtain the total information one, however, cannot perform the set of measurements successively because, unlike the classical case, the information obtained in successive quantum measurements depends on the order of its acquisition (see Fig. 5 and discussion above). Instead it seems that any attempt to obtain the total information content of a quantum system has to be related to the complete set of mutually complementary experiments performed on systems that are all in the same quantum state.

Concluding this discussion, we therefore suggest that it is natural to require that the total information content in a system in the case of quantum systems is the *sum* of the individual amounts of information over a complete set of m mutually complementary observables. As already mentioned above, for a spin-1/2 particle these are three spin projections along orthogonal directions. If we define the information gain in an individual measurement by the Shannon measure, the total information encoded in the three spin components is given by

$$H_{total} := H_1(p_x^+, p_x^-) + H_2(p_y^+, p_y^-) + H_3(p_z^+, p_z^-). \quad (14)$$

Here, e.g., p_x^+ is the probability to find a particle with spin up along the direction x.

Considering now an explicit example we will show that the total information H_{total} based on the Shannon measure is in general *not* invariant under unitary transformations. We calculate Eq. (14) for a spin-1/2 particle in the state $|\psi\rangle = \cos \theta/2 |z+\rangle + \sin \theta/2 |z-\rangle$ and we find that

$$H_{total} = -\frac{1-\sin\theta}{2}\log\frac{1-\sin\theta}{2} - \frac{1+\sin\theta}{2}\log\frac{1+\sin\theta}{2}$$
$$- \cos^2\frac{\theta}{2}\log\left(\cos^2\frac{\theta}{2}\right) - \sin^2\frac{\theta}{2}\log\left(\sin^2\frac{\theta}{2}\right) + 1 \quad (15)$$

depends on the parameter θ, thus being not invariant under unitary transformations. This associates a number of highly counterintuitive properties to H_{total}: (1) it can be different for states of the same purity (e.g., it takes its maximal value of 2 bits of information for $\theta=0$ and it takes its minimal value of 1.36 bits for $\theta=\pi/4$), (2) it changes in time even for a system completely isolated from the environment where no information can be exchanged with environment, and (3) it can take different values for different sets of the three orthogonal spin projections. These unnatural properties

---

[9]Consider a situation where instead of using single systems to send information to the receiver a sender uses a sequence of N systems where each individual system is drawn from an ensemble of pure states $\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$, with frequency of occurrence $\{w_1, \ldots, w_n\}$, respectively. It was shown in [18] that for sufficiently large N there are $2^{NS(\hat{\rho})}$ highly distinguishable sequences of pure states which become mutually orthogonal as $N \to \infty$. Here $S(\hat{\rho}) = -\text{Tr}(\hat{\rho} \log \hat{\rho})$ is the von Neumann entropy and $\hat{\rho} = \Sigma_i^n w_i|\psi_i\rangle\langle\psi_i|$. This means that if the sender uses a sequence consisting of a choice of states that respects the *a priori* frequencies w_i, and the receiver distinguishes whole sequences rather than individual states, then the (Shannon) information transmitted per system can be made arbitrarily close to $S(\hat{\rho})$. Here again the total density matrix $\hat{\rho}^N$ of N systems can be made arbitrarily close to the one for a classical mixture of the $2^{NS(\hat{\rho})}$ sequences of states.

[10]To specify a system described by a $n \times n$ density matrix completely one needs $n^2 - 1$ independent real numbers. Any individual, complete measurement (we consider here only complete measurements, i.e., where operators associated with the measurements are without degeneracy) with n possible outcomes defines $n-1$ independent probability values (the sum of all probabilities for all possible outcomes in an individual experiment is 1). Therefore, just on the basis of counting the number of independent variables, we expect that the number of different measurements we need in order to determine the density matrix completely is $(n^2-1)/(n-1)=n+1$. Ivanovic [19] and Wootters and Fields [20] demonstrated the existence of exactly $n+1$ mutually complementary observables by an explicit construction in the cases where n is a power of a prime number.

we see as another strong indication of the inadequacy of the Shannon measure to define the information gain in an individual quantum measurement.

V. SUGGESTED ALTERNATIVE MEASURE OF INFORMATION

Generalizing our above suggestion, we propose that it is natural to require that the information content of the quantum system defined as a sum of individual measures over a complete set of mutually complementary measurements is invariant under unitary transformations. Having shown that this cannot be achieved with the Shannon measure of information we now introduce a new measure of information that differs both mathematically and conceptually from Shannon's measure of information and according to which the information content has the required invariance property.

The new measure of information for an individual measurement with n possible outcomes is a quadratic function of probabilities[11]

$$I(p_1, \ldots, p_n) = \sum_{i=1}^{n} \left(p_i - \frac{1}{n} \right)^2, \qquad (17)$$

and it takes into account that for quantum systems the only features known before an experiment is performed are the probabilities for various events to occur [see [4] for discussion; there a specific normalization factor in expression (17) was used, resulting in maximally k bits for $n = 2^k$ possible outcomes]. This new measure fulfills Shannon's postulate 1 and is in agreement with 2, considering that the term $1/n$ in Eq. (17) is just an n-dependent zero offset. But it is in conflict with postulate 3, whose inadequacy we showed above. In particular, the measure $I(p_1, \ldots, p_n)$ is continuous in p_i and takes its maximal value of $(n-1)/n$ if one $p_i = 1$ and it takes its minimal value of 0 when all p_i are equal.

The important property of the new measure of information is that the total information defined with respect to it is *invariant* under unitary transformations. Using Eq. (13) one obtains that the sum over individual measures of information of mutually complementary observations results in [24]

$$I_{total} := \sum_{j=1}^{m} I(p_1^j, \ldots, p_n^j) = \sum_{j=1}^{m} \sum_{i=1}^{n} \left(p_i^j - \frac{1}{n} \right)^2 = \mathrm{Tr}\, \hat{\rho}^2 - \frac{1}{n}, \qquad (18)$$

---

[11]Expressions of the general type of Eq. (17) were studied in detail by Hardy, Littlewood, and Pólya [21]. They introduced a general class of mathematical expressions

$$M_\alpha = \left(\sum_{i=1}^{n} p_i^\alpha \right)^{\alpha-1} \qquad \text{for} \quad 0 \leq \alpha \leq \infty \qquad (16)$$

that from the point of view of information theory all can be assumed to quantify information properly. These expressions are also closely related to Tsallis's [22] nonextensive entropy $S_\alpha = [1/(1 - \alpha)] \Sigma_{i=1}^{n} (p_i^\alpha - 1)$ and Rényi's [23] entropy $H_\alpha = [1/(1 - \alpha)] \ln \Sigma_{i=1}^{n} p_i^\alpha$.

for a system described by the density matrix $\hat{\rho}$. Here p_i^j denotes the probability to observe the ith outcome of the jth observable. The total information content of the system therefore might all be encoded in one single observable or, alternatively, it might be partially encoded in all m mutually complementary observables. For a composite system in a product state the total information can all be encoded in individual systems constituting the composite system or, alternatively, in the extreme case of maximally entangled states it can all be encoded in joint properties of the systems with no information left in individual systems [4].

Independent of the various possibilities to encode information the total information content of the system cannot fundamentally exceed the maximal possible amount of information that can be encoded in an individual observable $[=(n-1)/n]$. This upper limit is reached when the system is in the pure state. When the system is in a completely mixed state the total information takes its minimal value of 0.

The property of invariance under unitary transformations implies that the total information content of the system does not depend on the particular set of mutually complementary observables chosen; it is a characterization of the state of the system alone, not of the specific reference set of complementary observables. Furthermore, since evolution in time is described by a unitary operation, the total information of the system is conserved in time if there is no information exchange with the environment.

We would like to note that the total information (18) was used in [25] to study the transfer of entanglement and information for quantum teleportation of an unknown entangled state through noisy quantum channels. The total information (18) belongs to the set of quantum counterparts of nonextensive entropies finding its application in an increasing number of problems in quantum physics, e.g., the description and controlling of laser cooling [26], a nonextensive approach to the decoherence problem [27], description and quantifying of entanglement, and deducing criteria for separability of density matrices [28,29].

VI. CONCLUSIONS

In this work we have stressed some conceptual difficulties arising when Shannon's notion of information is applied to define information gain in a quantum measurement. In particular we find that the axiomatic derivation of Shannon's measure of information is not applicable in quantum measurements in general. We also show that the information content of a quantum system defined according to Shannon's measure possesses some strongly nonphysical properties. We argue that these difficulties in defining the information gain in quantum measurement by the Shannon measure of information arise whenever it is not possible, not even *in principle*, to assume that attributes observed are assigned to the quantum system before the observation is performed.

Having criticed Shannon's measure of information as being not appropriate for identifying the information gain in quantum measurement we proposed a measure of information in quantum mechanics that both mathematically and

conceptually differs from Shannon's measure of information. While Shannon's information is applicable when the measurement reveals a preexisting property, our measure of information takes into account that for quantum systems the only features known before an experiment is performed are the probabilities for various events to occur. In general, which specific event occurs is objectively random.

The total information content of a quantum system defined according to our new measure of information as the sum of the individual measures of information for mutually complementary observations is invariant under unitary transformations. This implies that the total information content of the system is invariant under transformation from one complete set of complementary variables to another and is conserved in time if there is no information exchange with an environment.

[1] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948). A copy can be found at www.math.washington.edu/~hillman/Entropy/infcode.html

[2] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).

[3] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic Dordrecht, 1995).

[4] Č. Brukner and A. Zeilinger, Phys. Rev. Lett. **83**, 3354 (1999).

[5] A. Feinstein, *Foundation of Information Theory* (McGraw-Hill, New York, 1958), p. 17.

[6] J. Uffink, Ph.D. thesis, Utrecht University, 1990 (unpublished).

[7] This should be contrasted to the cases where the notion of information refers to knowledge about an unknown parameter in a probability distribution [R. A. Fisher, Proc. Cambridge Philos. Soc. **22**, 700 (1925), reprinted in R. A. Fisher, *Contributions to Mathematical Statistics* (Wiley, New York, 1950)], or the information for discriminating between two probability distributions [S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959)], or the information that one event provides about another event [I. M. Gelfand and A. M. Yaglom, in *Arbeiten zur Informationstheorie II*, edited by H. Grell (Deutscher Verlag der Wissenschaften, Berlin, 1957), p. 7. Russian original in Usp. Mat. Nauk **11**, 3 (1957)].

[8] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).

[9] D. K. Faddeev, in *Arbeiten zur Informationstheorie I*, edited by H. Grell (Deutscher Verlag der Wissenschaften, Berlin, 1957), p. 88. Russian original in Usp. Mat. Nauk **11**, 227 (1956).

[10] E. T. Jaynes, *Probability Theory: The Logic Of Science*. To our knowledge this book is only available on the web: http://bayes.wustl.edu/etj/prob.html

[11] E. Schrödinger, Naturwissenschaften **23**, 807 (1935); in *Quantum Theory and Measurement*, edited by J. A. Wheeler and W. H. Zurek (Princeton University Press, Princeton, 1983). A copy can be found at www.emr.hibu.no/lars/eng/cat

[12] W. Pauli, in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1958), Band Vol. 1, p. 7.

[13] J. S. Bell, Physics (Long Island City, N.Y.) **1**, 195 (1964).

[14] D. M. Greenberger, M. Horne, A. Shimony, and A. Zeilinger, Am. J. Phys. **58**, 1131 (1990).

[15] M. O. Scully, B. G. Englert, and H. Walther, Nature (London) **351**, 111 (1991).

[16] M. A. Horne, A. Shimony, and A. Zeilinger, Phys. Rev. Lett. **62**, 2209 (1989); T. J. Herzog, P. G. Kwiat, H. Weinfurther, and A. Zeilinger, *ibid.* **75**, 3034 (1995).

[17] E. T. Jaynes, in *Statistical Physics*, Brandeis Summer Institute, edited by K. W. Ford (Benjamin, New York, 1962).

[18] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996).

[19] I. Ivanovic, J. Phys. A **14**, 3241 (1981).

[20] W. K. Wootters and B. D. Fields, Ann. Phys. (N.Y.) **191**, 363 (1989).

[21] G. Hardy, J. E. Littlewood, and G. Pólya *Inequalities* (Cambridge University Press, Cambridge, England, 1952).

[22] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[23] A. Rényi, *Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie* (Deutscher Verlag der Wissenschaft, Berlin, 1962).

[24] Č. Brukner, Ph.D. thesis, Technical University of Vienna, 1999.

[25] J. Lee and M. S. Kim, Phys. Rev. Lett. **84**, 4236 (2000); J. Lee, M. S. Kim, Y. J. Park, and S. Lee, J. Mod. Opt. **47**, 2157 (2000).

[26] A. Bartana, R. Kosloff, and D. J. Tannor, J. Chem. Phys. **106**, 14 358 (1997); D. J. Tannor and A. Bartana, J. Phys. Chem. **103**, 10 359 (1999).

[27] A. Vidiella-Barrenco and H. Moya-Cessa, quant-ph/0002071.

[28] R. Horodecki and M. Horodecki, Phys. Rev. A **54**, 1838 (1996); R. Horodecki, P. Horodecki, and M. Horodecki, Phys. Lett. A **210**, 377 (1996).

[29] S. Abe and A. K. Rajagopal, Physica A **289**, 157 (2001); Chaos Solitons Fractals (to be published); C. Tsallis, S. Lloyd, and M. Barauger, e-print quant-ph/0007142.